

Al- Azhar University – Gaza

Faculty of Economics and Administrative Sciences

Department of Applied Statistics

Estimating Victimization Rates, Trends and Risk Factors in Palestine

تقدير معدلات واتجاهات وعوامل المخاطرة للإيذاء في فلسطين

By Nisreen Fouad Nour Aldin

A Thesis Submitted in Partial Fulfillment of Requirements for the

Degree of M.Sc. of Applied Statistics

Under the Supervision Of

Dr. Mahmoud K. Okasha

Associate Professor of Statistics

June , 2012



Al- Azhar University – Gaza

Deanship of Graduate Studies

Estimating Victimization Rates, Trends and Risk Factors in Palestine

A Thesis submitted in partial fulfillment of requirements for the Degree

of Master of Applied Statistic

by

Nisreen Fouad Nour Aldin

Committee of Evaluation

Title

Signature

1- Dr. Mahmoud K. Okasha Supervisor Associate Professor of Statistics

2- Dr. Abdullah Alhabil Internal Examiner Associate Professor of Statistics

3- Dr. Hazem I. ELShekh Ahmed

External Examiner Assistant Professor of Statistics

Faculty of Economics and Administrative Sciences

Department of Applied Statistics

Gaza – Palestine 2012

Estimating Victimization Rates, Trends and Risk Factors in Palestine

ٳڰڟٳٵ

إلى والدتي العزيزة ووالدي الناضل

إلى مرمح زوجي الراحل

إلى أطفالي الأعزا

إلى أصدقائي المخلصين

إلى كل من وقف بجانبي وسانديني

اهدي هذا الجهود المنواضع .

ACKNOWLEDGMENT

There are many people I would like to acknowledge for their assistance and support during my elicitation journey.

Thanks to Dr. Mahmoud K. Okasha, my supervisor for his encouragement and advice . I am forever grateful for your guidance and assistance on this dissertation.

I am grateful to Dr. Abdullah alhabil, and Dr. Hazem ELShekh Ahmed for their efforts, notes and advices.

Lastly, I would like to thank the Palestinian Central Bureau of Statistics (PCBS) for providing me with the necessary data to complete this dissertation.

DECLARATION

I certify that this thesis submitted for the Master degree is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any university or institution.

Signed

Nisreen Fouad Nour Aldin

Abstract

The current study examined different risk factors that are most informative in identifying victims at the greatest risk to be victimization. These factors include region, sex, job, owning a car, which party prone to tangible losses, place of the crime, attempting break the house and reception a threat call.

Logistic regression model has been used to identify the most information factors. The effects of each factor and its odds ratio and risk ratio are examined.

Results of a logistic regression analysis revealed the fact that the most influenced factors on victimization according to the final model consist of the following two risk factors: sex and owning a car.

The model has been applied to predict the occurrence of persons to be victimized and succeeded in correctly predicting (64%) of people who have really fallen victims and (96%) of people who are unbeatable to crime. The general percentage of correct prediction was (93.5%).

vii

ملخص باللغة العربية

اهتمت الدراسة الحالية بالبحث في العوامل التالية باعتبار ها أهم العوامل المفيدة في التعرف على ضحايا الجرائم وقد تمثلت في : المنطقة، الجنس، الوظيفة، امتلاك سيارة، الجهة التي تحملت الخسائر المادية، مكان حدوث الجريمة، محاولة اقتحام البيت و استقبال مكالمة تهديد. ولقد استخدم نموذج الانحدار اللوجستي لتحديد أهم عوامل الخطر بالإضافة الى اختبار تأثيرات كل عامل ونسب الخلاف . وقد كشفت نتائج تحليل الانحدار اللوجستي استنادا إلى النموذج النهائي أن هناك عاملين أكثر

وقد تم تطبيق النموذج للتنبؤ بالضحايا ونجح في التنبؤ بشكل صحيح بنسبة (64٪) من الاشخاص الذين وقعوا فعلا ضحايا لجريمة ما ونسبة(96٪) من الناس الذين هم عرضة للجريمة بشكل عام كانت نسبة التنبؤ الصحيح (93.5٪).

Table of Contents

Content	No.
Chapter 1 : Introduction	1
1.1 Background	1
1.2 The data	3
1.3 Research problem	4
1.4 Research importance	4
1.5 Methodology	5
1.6 Thesis Structure	5
1.7 Historical review	6
Chapter 2: Concepts, Definitions and The Sample	17
2.1 Concepts and Definitions	17
2.2 The sample	19
2.3 Data processing	21
2.4 Data Quality	21
2.5 The Questionnaire	23
2.6 The salient features of the data	24
Chapter 3: Logistic Regression	28
3.1 Interprets the logistic regression	28
3.2 Linear Approximation Interpretation	29
3.3 Logistic regression with retrospective studies	31
3.4 Inference for logistic regression	32
3.5 Building logistic regression models	36
3.6 AIC, Model selection, and the "Correct" Model	39

3.7 Summarizing predictive power: Classification tables	40
3.8 Model checking	41
Chapter 4 : Data analyses	43
4.1 Introduction	43
4.2 Data description	44
4.3 Logistic regression analyses	51
4.4 Odds ratio	53
Chapter 5 : Conclusion and Recommendations	56
5.1 Conclusion	56
5.2 Recommendations	56
References	58

List of tables

Table	No.
Table 2.1: Main indicators of victimization survey	28
Table 3.1: Logits implied by indicator variables in model	35
Table 4.1: Summary of logistic regression analysis results	52
Table 4.2: Accuracy of predictions of victim versus no victim: Final	55
Models	

List of Figures

Figure	No.	
Figure 3.1 : Logistic regression function	29	
Figure 3.2 : Linear approximation to logistic regression curve	30	
Figure 4.1: Percentage distribution of victimized persons in	45	
Palestinian Territory by last criminal offense		
Figure 4.2: Percentage distribution of victimized persons in	46	
Palestinian Territory by Location of last crime and type of Locality		
Figure 4.3: Percentage of Victimized Persons by Last crime reporting and Region	47	
Figure 4.4: Percentage of Victimized Persons by Perpetrator of last criminal		
offense and Region		
Figure 4.5: Percentage of Victimized Persons by Party prone to tangible losses of		
the last criminal offense and Region		

List of Abbreviations

Abbreviation	Stands for
ACBS	Palestinian Central Bureau of Statistics
HLM	Hierarchical Linear Models
AIC	Akaike Information Criterion
PSUs	Primary Sampling Units
EAs	All Enumeration Areas
H-L	Hosmer-Lemeshow Test
PAC	Percentage Accuracy in Classification
VIF	Variance Inflation Factor
ML	Maximum Likelihood

Chapter 1

Introduction

1.1 Background and Aims

The aim of this thesis is to understand the determination of victimization rates in Palestine, by applying various empirical analyses. The concept of victim dates back to ancient cultures and civilizations. Over the centuries, the word *victim* came to have additional meanings. During the founding of victimology in the 1940s, victimologists such as Mendelson (1963) and Von Hentig (1948) tended to use textbook or dictionary definitions of victims as hapless dupes who instigated their own victimizations. This notion of "victim precipitation" was vigorously attacked by feminists in the 1980s, and was replaced by the notion of victims as anyone caught up in an asymmetric relationship or situation. "Asymmetry" means anything unbalanced, exploitative, parasitical, oppressive, destructive, alienating, or having inherent suffering. In this view, victimology is all about power differentials. Today, the concept of victim includes any person who experiences injury, loss, or hardship due to any cause. Also today, the word *victim* is used rather indiscriminately; e.g., cancer victims, accident victims, victims of injustice, hurricane victims, crime victims, and others. The thing that all these usages have in

common is an image of someone who has suffered injury and harm by forces beyond his or her control.

The term "crime victim" generally refers to any person, group, or entity who has suffered injury or loss due to illegal activity. The harm can be physical, psychological, or economic. The legal definition of "victim" typically includes the following:

• A person who has suffered direct, or threatened, physical, emotional or pecuniary harm as a result of the commission of a crime; or in the case of a victim being an institutional entity, any of the same harms by an individual or authorized representative of another entity. Group harms are normally covered under civil and constitutional law, with "hate crime" being an emerging criminal law development, although criminal law tends to treat all cases as individualized. Many victims feel that defining themselves as a "victim" has negative connotations, and choose instead to define themselves as a "survivor." This is a very personal choice that can only be made by the person victimized. The term "survivor" has multiple meanings; e.g. survivor of a crime, "survivor benefits". It remains to be seen whether this terminology for victims of crime will endure.

"Victim defenses" have recently emerged in cases of parricide (killing one's parents) and homicide of batterers by abused spouses. Advocates for

battered women were among the first to recognize the issue, and promote the "battered woman syndrome" to defend women who killed or seriously injured a spouse or partner after enduring years of physical, emotional and/or sexual abuse. Attorneys have also drawn upon theories of Posttraumatic Stress Disorder to defend their client's behavior. From time to time, media attention to these defenses becomes intense, and certain "high profile" cases tend to influence public opinion and spread confusion over who is the "victim" and who is the "victimizer."

Available official statistics on the Palestinian society in the Palestinian Territory lack adequate data pertaining to victimization. This situation has prompted the Palestinian Central Bureau of Statistics (PCBS) to conduct a fourth household survey on this subject, making it possible to describe and assess certain aspects of victimization focusing on households victimized by criminal offenses.

1.2 The data

The data of this study has been conducted by PCBS(2008) and based on a household sample survey during the period from 04/10/2008 until 31/12/2008. It provides basic indicators on various aspects of victimization, including households victimized by criminal offenses, type of criminal offense, tangible losses of crimes, Crime location, Perpetrator, Crime

reporting, Reported crime underwent legal proceedings, Who party prone to tangible losses. A special questionnaire was designed and recommendations in the field of victimization statistics while taking the Palestinian particularities into account. The questionnaire covers the following items:

- 1. Type of criminal offense
- 2. Crime location
- 3. Crime reporting
- 4. Perpetrator

1.3 Research problem

The victim is one of crime action pillars which consists of criminal, crime and the victim, it has a role in specification and shaping the criminal act. The problem to be studied in this thesis is to find out any existing specific characteristics of the victim in the Palestine society.

1.4 Research Importance

The importance of this study lies in little and scarcity of researches that handled the victim in Palestine society, the knowledge and situation of the victims of crimes which may have a role in being a victim of crime in order to make scientific and practical use to find the suitable solution, preventive and treatment plans to limit the spread and increase in victims and control crime causes.

1.5 Methodology

Since we have many independent binary variables and the response variable of interest in the survey such as "households victimized by criminal offense" is binary, the main statistical model used in the analysis of our victimization data set in this thesis is the logistic regression model . Logistic regression is an ideal model for analysis when a researcher needs to determine the best subset of independent variables, among various independent variables, that best predict group membership of cases using data of known groups of a dichotomous dependent outcome variable.

1.6 Thesis Structure

The thesis divided as follows

In chapter two, we provide the main concepts, definitions and the sample, including definitions for assault, crime, crime location, household, perpetrator, physical harm, robbery, theft and victim. This chapter also presents data processing, data quality, the questionnaire and the salient features of the data.

The focus of Chapter three is on the interprets the logistic regression, logistic regression with retrospective studies, inference for logistic regression, building logistic regression models, AIC, model selection, and the "correct" model, classification tables and model checking.

Chapter four focuses on data description, logistic regression analyses, odds ratio. Chapter five presents the conclusion and recommendations.

1.7 Historical review

In a study of *Rachel (2009)*, Six group of countries where highlights the importance of distinguishing between areas of the world beyond only industrial and non-industrial categorizations, during the four years: 1989, 1992, 1996, and 2000. The sample size for each country ranges from approximately 1,000 to 2,000 respondents per survey. Multilevel Equations by employing multilevel models, it is possible to examine individual level measures, random effects, structural level measures, and cross-level interactions. Hierarchical Linear Models (HLM) take into consideration the layered or nested nature of the data, nesting respondents within the country in which they reside. Within this research, the dependent variables are dichotomous, scored 1 if the respondent has been victimized and scored 0 if not.

Within the multilevel analysis, routine activities and lifestyle variables are included at the individual level of analysis. A total of six models are run for both assault and burglary victimization. All individual level variables are included within each model, highlighting the effect of respondents' routines and lifestyles on victimization experience. *Assault Victimization* of the variables measuring the routines of individuals, going out in the evening for leisure activities reaches significance, while whether one works or goes to school does not reach a significant level. A one unit increase in how often individuals go out for leisure activities results in a

7% increase in odds of assault victimization. Several of the risk factors increase the odds of victimization, including living alone (OR=1.38), living in an apartment (1.23), and being male (1.48). With a one unit increase in the age category, there is a decrease in the odds of being assaulted (OR=0.78).

Burglary Victimization the routine activities of individuals are important in burglary victimization risk. A one unit increase in how often respondents go out in the evening for leisure activities increases the odds of burglary victimization by 4%. Respondents who work or go to school have a 10% increase in odds of burglary victimization compared to those who do not work or go to school. Individuals who live in an apartment building as compared to a detached home have a 23% decrease in odds of property victimization. The risk factor of education increases the odds of burglary victimization (OR=1.02).

Marcotte and Markowit (2009) in their paper explored the relationship between trends in treatment for mental illness and violent crime. The researchers have tried to characterize the behavioral mechanisms for these relationships by summarizing important syndromes and how they might contribute to behaviors leading to criminal acts and also increase risk of victimization.

The researchers provide evidence that increased prescriptions for mental illness drugs in general are associated with decreases in violent crime. The researchers estimate that a one percent increase in the total prescription rate is associated with a 0.051 percent decrease in violent crimes. To put this in perspective, doubling the prescription rate would reduce violent crimes by 5 percent, or by about 27 crimes per 100,000, at the average rate of 518 crimes per 100,000 population. While doubling the prescription rate seems like a large change, it has been estimated that 28 percent of the U.S. adult population in any year has a diagnosable mental or addictive disorder, yet only 8 percent seeks treatment. Doubling the treatment rate would still leave a substantial portion of the ill untreated.

The small elasticities they estimate may of course be due to limited behavioral response to new therapeutic agents. However, even if the impact of treatment were substantial, effects can be hard to identify in community-based data like ours. A substantial limitation in population level data is that we do not know if treatment is going to those at risk for criminal behavior. There is obvious reason to be concerned that treatment is most available for those who otherwise have few risk factors for engaging in criminal violence.

Tyler et al.(2004) examined the risk factors associated with the likelihood of being sexually victimized by a stranger friend/acquaintance since being on the street was examined among 372 homeless and runaway youth. Young people were interviewed on the streets and in shelters by outreach workers using a systematic sampling strategy. Youth who engaged in more

high-risk behaviors were expected to be at greater risk for sexual victimization by both known and unknown assailants. Results indicated that for females, running from home for the first time at an earlier age was associated with sexual victimization by both a stranger and friend/acquaintance. However, engaging in deviant subsistence strategies, survival sex, and grooming predicted being sexually victimized by a friend/acquaintance. For males, survival sex and grooming predicted stranger sexual victimization, whereas sexual orientation was associated with sexual victimization by a friend/acquaintance. Overall, 35% of the sample had been sexually victimized.

Koo and Pierre (2003) conducted a study to achieve the following goals:

(1) Estimate the prevalence of violent victimization in a 30 days period among a sample of 900 street recruited heroin users in Miami- Dade County, Florida;

(2) Identify the risk factors for violent victimization among this drug group;

(3) Examine two different types of violent victimization (robbed and injured) and analyze whether risk factors vary among these different types of victimization.

Previous studies find that involvement in a deviant lifestyle increases the risk of victimization. Based on the assumptions of routine activities theory,

heroin addicts are at an increased risk of victimization due to their lifestyle activities. Thus, the lifestyle and context of the drug/street addict subculture is important to focus on when exploring the patterns of violent victimization.

The combination of these two theoretical perspectives creates a single conceptual framework that focuses on four domains:

_ Sociodemographics

_ Drug Use History

_ Lifestyle

_ Social Networks

Logistic regression was employed to examine the overall violent victimization and the different types of victimization by the four domains. Each domain was analyzed separately using logistic regression. Those independent variables at a significance level of .10 or less within their domain were entered into a final multiple logistic regression model.

Roodman (2000) conducted another study to examine two competing models of sexual victimization that examined the path between child abuse and later sexual victimization. Structural equation modeling was used to examine two competing models of sexual victimization. A sample of 276 college students taking introductory psychology were participants. They anonymously completed a packet of questionnaires that provided the indicator variables for the path models that were tested. Both models

tested demonstrated significant pathways between the factor for child abuse (comprising physical and sexual abuse) and negative cognitive schemas and for child abuse and dissociation. However, the paths from negative cognitive schemas and dissociation to sexual victimization (comprising both adolescent and adult sexual victimization) were not significant suggesting that, although these factors are influenced by child abuse, they do not mediate revictimization. Risky behaviors, as measured by consensual sex and alcohol consumption, do not appear to be influenced by early abuse, but there was a significant pathway between this factor and sexual victimization suggesting that these risky behaviors are independent risk factors for sexual victimization in adolescence and adulthood. In one model there was a significant pathway between child abuse and sexual victimization which is what would be expected if given previous findings that suggest past abuse is the best predictor of future victimization experiences. That the other model did not demonstrate this relationship was surprising.

Gaviria and Pagés (1999) in their paper they used the Latinobarometer to study the patterns of crime victimization in Latin America. The Latinobarometer is a public opinion survey covering 17 Latin American countries. The survey has been regularly conducted since 1996. Roughly, 1,500 individuals have been interviewed in each country each year. The sampling method varies slightly from country to country. The levels of victimization are staggering. In five countries (Peru, Ecuador, Mexico Venezuela, El Salvador and Guatemala) more than 40 percent of the urban households have had at least one member victimized during the previous year. In Guatemala at least one individual of every two households has been victimized. Spain, the only industrialized country included in the survey, exhibits the lowest victimization rates in the sample. Uruguay, Panama and Chile exhibit the lowest victimization rates in Latin America. An important shortcoming of the Latinobarometer is the absence of information about type of victimization. They will assume throughout that the victimization rates obtained from this survey correspond mainly to property crime --an assumption justified by the fact that violent crimes usually represent a small fraction of all crimes.

In addition to the Latinobarometer, the researchers used victimization surveys for three countries: Colombia, El Salvador and Peru. These surveys permit to refine the analysis in several respects. This survey covers only Metropolitan Lima and includes 8,643 individuals distributed in 2,473 households. The survey was conducted in 1997 and has six different modules, each dealing with a different type of crime.

Four modules apply to individuals (robberies, car thefts, assaults and vandalism) and two apply to households (burglaries and kidnappings). the researchers use information from all modules to compute victimization

rates at the household level in order to ensure comparability with the rates obtained from the Latinobarometer.

The victimization rates obtained from Lima and El Salvador surveys are very high. In Lima, more than 70 percent of the households had at least one member victimized in 1997. Robberies are the most common offense, followed by burglaries and assaults. In El Salvador, almost 60 percent of urban households suffered from some form of victimization. Also here property crimes constitute the bulk of all offenses. In Colombia, victimization rates are surprisingly small, only 12 percent of the household reported an incident during the past year. However, the small rates obtained for robberies and assaults seem to imply that only serious offenses were reported, reducing the comparability with the other surveys. Sampson (1985) uses National Crime Survey victimization data from 1973-1978 to examine the effects of neighborhood characteristics and extent of urbanization on rates of theft and violent personal victimization. The results underscore the importance of urbanization and the physical environment in predicting victimization risk. Regardless of age, racial composition, and poverty, both the extent of urbanization and housing density had significant positive effects on victimization Several important interactions were also uncovered Poverty tends to increase victimization risk only in urban areas, while density exerts an increased effect on victimization in suburban and rural areas. Overall the results confirm the

need for researchers to take into account both neighborhood factors and the urban-rural dimension in explaining victimization.

The higher level of crime in U.S. cities compared to suburban or rural areas has long been an accepted fact in criminology. Regardless of the data source used, crime statistics consistently show that urban crime rates are substantially greater than non-urban crime rates. Moreover, population size has been shown to be an important predictor of crime rates across U,B cities.' The apparent strong impact of urbanism on crime has led to an interesting development in the ecological study of *crime*: almost every ecological study of crime to date has been conducted within central cities. Both recent crime trends and theoretical developments suggest that the limitation of areal studies of crime to cities is unwarranted. First, crime trends indicate that mime is rising at a faster rate in suburban and rural areas than in large cities. Particularly noteworthy is the fact that rural crime rates today are roughly equal to urban crime rates of 1967. Clearly, the phenomenon of serious crime in the United States is not limited to the confines of our major cities. Indeed, the reality of criminal victimization is becoming commonplace in areas once considered idyllic setting. Thus, although the absolute level of crime is still higher in cities than in surrounding areas, projections from recent trends suggest that rate differences are quickly converging.

Acad Med (1986) focused on criminal homicide, defined as death due to injuries illegally inflicted by another person with intent to injure or kill by any means. Determination that a death was criminal homicide was based on the results of investigation by the Los Angeles Police Department.

Demographic characteristics of victims and perpetrators and situational characteristics of the homicide were obtained from confidential police files. In this study Hispanics are defined as people of Spanish descent. In Los Angeles this group includes not only Mexican-Americans, but also substantial numbers of immigrants from Central America and other locations. Anglo refers to non-Hispanic people who are white, and black refers to non-Hispanic people who are black. Population data used to calculate rates in this report were generated by linear interpolation between the 1970 and 1980 published and unpublished census data for the population of the city of Los Angeles by age, race/ethnicity, and sex. Results of toxicologic analyses performed on the blood or tissues of homicide victims were abstracted from the files of the Los Angeles Medical Examiner-Coroner and linked to data obtained from police files. From 1970-79 only blood alcohol and barbiturate levels were routinely determined for homicide victims. Therefore, he limited the discussion of the results to alcohol and barbiturate use by homicide victims. Demographic patterns in the risk of homicide victimization.

During the period 1970-79 criminal homicide took a total of 4,950 lives in the city, a 10-year rate of 17.1 homicides per 100,000 population. Homicide took its greatest toll among men, the young and minorities. Men were almost four times more likely to become homicide victims than were women.

Age-specific rates peaked at 26.9 per 100,000 population in the 25- to 34year-old group. Blacks were at greatest risk of victimization, with a rate of 45.6 per 100,000 population. For blacks and Hispanics the risk of homicide victimization was 5.6 and 2.3 times greater, respectively, than that for Anglos.

When examined by race/ethnic group and sex, the risk of homicide victimization was greatest for black men, followed by Hispanic males and by black women. The risk of homicide victimization for males relative to females was much greater for Hispanics than for blacks and Anglos.

Hispanic men were at 7.3 times greater risk than Hispanic women, while black and Anglo men were at 4.3 and 2.3 times greater risk than black and Anglo women respectively. Relative differences in race/sex-specific homicide rates were unchanged after rates were age-adjusted.

Black men were at greatest risk of homicide victimization in every age group. Rates for Hispanic men peaked at an earlier age than those for black or Anglo men. Rates for Anglo men were fairly uniform across the age groups.

CHAPTER 2 Concepts, Definitions and The Sample

2.1 Concepts and Definitions

In this section we provide some basic definitions of the legal and social concepts that will be used in our study. It is important to note that the survey that we are using in this study has been conducted by the Palestinian Central Bureau Of Statistics in 2008. It covers all the Palestinian Territories including the West Bank and Gaza Strip. It is composed of 10260 sampling units.

2.1.1 Assault:

The term assault refers to physical attack against persons, but excludes indecent assault. Some criminal or penal codes distinguish between aggravated and simple assault depending on the degree of resulting injuries.

2.1.2 **Crime:**

Crime is defined as any act involving violation of laws or public rights duties towards the state or society in general.

2.1.3 Crime Location:

This term refers to the place where the crime took place.

2.1.4 Household:

It refers to a group of at least one person living together who make common provisions for food or other essentials for living. Households members may be related, unrelated or a combination of both.

2.1.5 Number of Households (n):

Since the household is the sampling unit, the number of households is the sample size . Sometimes, it is weighted sample size .

2.1.6 Perpetrator:

The person violating effective laws by undertaking criminal events against other persons or their properties is referred to as perpetrator.

2.1.7 Physical Harm:

All losses that a person may suffer during the crime that took place in the last 12 months prior to the survey, which resulted in wounds, murder, malformation or disability is referred to as physical harm.

2.1.8 Properties:

All movable and fixed assets belonging to the individuals (household members) regardless of whether they are inside or outside the house are termed in this study as properties.

2.1.9 Robbery:

Illegally breaking into the property of somebody with the intention to commit a crime is referred to as robbery.

2.1.10 Theft:

The term theft in this study refers to the removal of property without the property owner's consent. Theft excludes burglary and house breaking; it includes the theft of a motor vehicle, shoplifting and other minor offenses, e.g. pilfering and petty theft may or may not be included as thefts.

2.1.11 Victim:

The person affected by an offense or loss or prey to catastrophic, criminal or brutal events is termed here as victim. Any person who was offended and whose properties were partially or totally affected by a criminal act or incident is classified as a victim.

2.2 The Sample

The survey's methodology was designed taking into account the Palestinian conditions, international standards, data processing requirements and the comparability of outputs with other related surveys conducted in the Palestinian Territory for 2008. The sample is composed of 10260 sampling units for all Palestinian governments .

2.2.1 Sampling Frame

The sampling frame consisted of a master of all enumeration areas (EAs) selected from the population housing and establishment census of 2007. The sampling frame consists of area units of relatively equal size (the number of households in each unit is about 150 housing units), and these units have been used as primary sampling units (PSUs).

2.2.2 Sample Design

The sample is a two-stage stratified cluster systematic random sample. The sample of this study was applied to all households in round 51 of the labor force survey.

2.2.3 Stratification:

Two levels of stratification were made:

1. Stratification by Governorates (16 Governorates)

2. Stratification by type of Locality which comprises:

(a) Urban (b) Rural (c) Refugee Camps

2.2.4 Sample Size:

The sample size equals 10,263 households. The sample is distributed over 491 enumeration areas in the West Bank and Gaza Strip.

Sample design considered the target cluster size or "sample-take," the number of households to be selected per PSU on the average. In this survey 10,263 households has been selected from 491 master sample areas in all Palestinian governorates.

2.2.5 Weighting

Weights have been calculated for each sampling units. Weights reflect the sampling procedures. Adjusted weight is important to reduce bias resulting from non-responses.

2.3 Data Processing

Both data entry and tabulation were conducted using the ACCESS and SPSS software programs. Data entry has been organized into two files, corresponding to the main parts of the questionnaire. Data entry template has been designed to reflect an exact image of the questionnaire, and included various electronic : logical, range, consistency and crossvalidation checks.

2.4 Data Quality

It is very important to calculate standard errors for the main survey estimates, so that we can identify the accuracy of estimates and the survey reliability. Errors of the survey are of two kinds: statistical errors, and nonstatistical errors. Non-statistical errors are related to the procedures of statistical work at different stages, such as the failure to explain questions in the questionnaire, unwillingness or inability to provide correct responses, low statistical coverage, etc. These errors depend on the nature of the work, training, supervision, and conducting of all the various related activities. However, it is difficult to estimate numerically such errors due to absence of technical computation methods based on theoretical principles to tackle them.

On the other hand, statistical errors can be measured by the standard error, which is the positive square root of the variance. The variance of this survey has been computed by using SPSS package.

Sampling rather than comprehensive enumeration has been used to collect data in this survey. Therefore it is liable to two types of errors affecting the quality of survey data, sampling (statistical errors) and non-sampling errors (non-statistical errors). Statistical errors mean the errors resulting from sample designing and this is computed simply. Variance and effect of sample design has been computed for the Palestinian Territory, the West Bank and Gaza Strip.

Non-statistical errors, could not be determined easily, due to the diversity of sources from which they may arise, e.g., the interviewer, respondent, editor, coder, and data entry operator.

However, several measures were adopted to minimize the effects of nonstatistical errors on the data. The data entry program was programmed in a way that allows error detection and correction, particularly logical errors that might not be discovered before data entry. Consistency check was applied to assure accuracy after data entry.

There are different methods to evaluate data according to subjects, and they include:

• Frequency of missing values and responses like "other" or "do not know" and examining data consistency between the different sections.

• Comparing survey results with other sources; and also, with results of Victimization Survey 1996, 1999, all utilized quality checks revealed that data of this survey is of a high quality.

2.5 The Questionnaire

Available official statistics on the Palestinian society in the Palestinian Territory lack adequate data pertaining to victimization. This situation has prompted the PCBS to conduct a fourth household survey on this subject, making it possible to describe and assess certain aspects of victimization focusing on households victimized by criminal offenses.

This study is based on a household sample survey conducted during the period from 04/10/2008 until 31/12/2008. It provides basic indicators on various aspects of victimization, including households victimized by criminal offenses, type of criminal offense, tangible losses of crimes, etc. A special questionnaire was designed in accordance with UN standards and recommendations in the field of victimization statistics while taking the Palestinian particularities into account. The questionnaire covers the following items:

- 1. Type of criminal offense
- 2. Crime location
- 3. Crime reporting
- 4. Perpetrator
2.6 The salient features of the data

2.6.1 Victims of Criminal Offenses at the Households Level

The results showed that (7.5%) of the Palestinian households in the Palestinian Territory were exposed to criminal offenses: (5.8%) of West Bank households were victimized and (10.9%) in Gaza Strip.

The results showed that the percentage of households in the Palestinian Territory that were exposed to theft (excluding vehicles theft) was (2.0%), vehicle theft or part of it was (6.1%), property damage was (1.7%), threat was (0.8%), and assault was(0.9%).

General data revealed that the percentage of households victimized by criminal offenses in Gaza Strip is higher than that in West Bank, except for households victimized by vehicle theft or part of it): (6.2%) in West Bank and (5.9%) in Gaza Strip, robbery or theft attempt: (0.7%) in West Bank and (0.3%) in Gaza Strip.

The results showed that (2.4%) of households in the Palestinian Territory were exposed to Israeli Soldiers or Settlers Harassment and Assault, compared with (7.1%) in 2004.

2.6.2 Victims of Criminal Offenses at the Individual Level

Type of Criminal Offense

The results showed that (33.9%) of individuals victims of criminal offenses in the Palestinian Territory were exposed to theft and attempted theft; (42.1%) of persons in West Bank and (26.3%) of persons in Gaza

Strip; threat\ assault occurred to (18.4%); (12.3%) in the West Bank and (24.0%) in Gaza Strip, exposed to property damage (18.3%); (13.3%) in West Bank and (23.0%) in Gaza Strip, and Israeli Soldiers or Settlers Harassment and Assault affected (27.8%) of victimized persons; (30.1%) in West Bank and (25.6%) in Gaza Strip.

Crime Location

The results indicated that (45.7%) of criminal offenses in the Palestinian Territory took place inside the house, (26.5%) nearby the house, (18.7%) took place in another place inside locality, and (9.1%) outside the locality. Criminal offenses occurring inside the house were higher in Gaza Strip (52.2%) than in the West Bank (38.8%), and criminal offenses that took place outside the locality were lower in Gaza Strip (3.7%) compared with the West Bank (14.8%).

Crime Reporting

The results showed that (53.0%) of persons victimized by criminal offenses in the Palestinian Territory reported the crimes, (47.0%) in West Bank and (57.6%) in Gaza Strip. The results showed that (14.2%) of crime reporting underwent legal proceedings.

Perpetrator

The results showed that (33.2%) of criminal offenses against persons in the Palestinian Territory were committed by Israeli Soldiers or Settlers, (41.7%) in West Bank and (25.1%) in Gaza Strip. About (3.8%) of these criminal offenses were committed by one of the relatives, (6.5%) in West Bank and (1.4%) in Gaza Strip.

Physical Harm and Tangible Losses of Criminal Offenses

The percentage of criminal offenses that caused physical harm was (12.2%) in the Palestinian Territory, in West Bank (13.1%) compared with the Gaza Strip (11.5%).

The results show that (30.4%) of criminal offenses against persons in the Palestinian Territory caused tangible losses for more than 1000 Jordanian Dinars, distributed as (34.2%) in West Bank and (26.9%) in Gaza Strip. In about 72.7 of criminal offenses cases against persons in the Palestinian Territory, the victim was subjected to tangible losses, compared with (85.8%), (88.7%) and (78.6%) in the years 1996, 1999 and 2008 respectively.

Table 2.1:Main Indicators of Victimization Survey - 1996, 1999, 2004, 2008

Indicator	1996	1999	2004	2008
Victims of Criminal Offenses at the Household Level			<u></u>	
Descenters of vistimized households of all ariminal offeness	5.0	F 4	44.0	76
Percentage of victimized households of all criminal offenses	5.0	5. 1	11.3	7.5
Percentage of households exposed to their (excluding vehicle)	1.6	1.2	1.2	2.0
Percentage of households exposed to vehicle their of part of it	1.0	1.9	1.1	0.1
Percentage of households exposed to robbery of their attempt		0.5	0.4	0.0
Percentage of households exposed to property damage	1.3	0.2	1.5	1.7
Percentage of households exposed to infeat	1.3	0.4	1.5	0.0
Percentage of households exposed to lerabli soldiers or settlers	1.2	0.4	1.5	0.9
harasement or assault		13	71	24
Percentage of households exposed to other crimes		1.5	0.3	0.4
			0.0	0.4
** Victims at the Individual Level by Last Criminal Offense				
Percentage of persons exposed to theft theft robbery attempt	54.2	55.2	19.5	33.9
Percentage of persons exposed to threat assault	18.8	18.0	13.1	18.4
Percentage of persons exposed to property damage	16.1	4.4	8.2	
Percentage of persons exposed to Israeli soldiers and settlers harassment				
or assault			56.6	
Location of Last Crime				
Percentage of persons exposed to criminal offense inside house	23.5	16 5	44.8	45 7
Percentage of persons exposed to criminal offense nearby house	32.8	41 7	17.5	26.5
Last Crime Reporting	52.0	71.7	17.5	20.5
Percentage of victimized persons who reported the crime	40.2	43.2	29.5	53.0
Reasons for Not Reporting Last Crime				
Percentage of victimized persons not reporting because crime not serious				
Enough		51.7	29.3	28.5
Percentage of victimized persons not reporting because personal\ tribal				
solution		10.7	30.0	18.3
Percentage of victimized persons not reporting because preferring no				
interference of police		10.7	20.7	21.2
Perpetrator of Last Crime				
Percentage of persons exposed to criminal offense from Israeli soldiers or				
settlers	11.6	26.8	62.7	33.2
Percentage of persons exposed to criminal offense from a relative	13.0	8.9	4.4	3.8
Physical Harm and Tangible Losses of Last Crime				
Percentage of percens exposed to criminal offense and caused physical				
harm	22.6	16.5	10.5	12.2
Percentage of percens exposed to criminal offense and caused tangible	22.0	10.5	10.5	12.2
	77 /	61.0	60.0	64.0
Percentage of persons exposed to criminal offense and caused physical	77.4	01.0	00.0	04.0
harm and tangible losses		3.2	34	1.8
Percentage of persons exposed to criminal offense and caused tangible		0.2	0.1	
losses of more than 1000 Jordanian Dinars	15.3	14.7	21.2	30.4
Party Prone to Tangible Losses of Last Crime				
Percentage of persons exposed to criminal offense and the victim was				
prone to tangible losses	85.8	88.7	78.6	72.7

(-): Means data not available. **1996, 1999 the definition of criminal offenses includes: (theft, assault, property damage and other crimes), while in 2004 the definition of criminal offenses includes: (theft, robbery or theft attempt, threat, assault and property damage, Israeli soldiers and settlers harassment or assault and other crimes)

CHAPTER 3

Logistic Regression Model

Since we have many binary response variables of interest in the survey such as Households victimized by criminal offense, the main statistical model used in the analysis in this thesis is the logistic regression model. In this chapter we are going to discuss this model in some detail.

Let us now take a closer look at the statistical modeling of binary response variables, for which the response outcome for each subject is a "success" or "failure". The most popular model for binary data is *logistic*

regression.

3.1 Interprets the logistic regression model

To begin, suppose there is a single explanatory variable X, which is quantitative. For a binary response variable Y, recall that $\pi(x)$ denotes the "success" probability at value x. This probability is the parameter for the binomial distribution. The logistic regression model has linear form for the *logit* of this probability,

$$\operatorname{logit}[\pi(x)] = \operatorname{log}(\pi(x) / 1 - \pi(x)) = \alpha + \beta x \tag{1}$$

The formula implies that $\pi(x)$ increases or decreases as an S-shaped function of *x* (Figure 3.1).



Figure 3.1 : Logistic regression functions

The logistic regression implies the following formula for the probability $\pi(x)$,

$$\pi(x) = \exp(\alpha + \beta x)/1 + \exp(\alpha + \beta x)$$
(2)

This section shows ways of interpreting these model formulas.

3.2 Linear Approximation Interpretations

The logistic regression formula (1) indicates that the logit increases by β for every 1 unit increase in *x*. The interpretation of the logit scale is soretires misleading, so we need to consider alternative interpretations.

The parameter β in equations (1) and (2) determines the rate of increase or decrease of the S-shaped curve for $\pi(x)$. The sign of β indicates whether the curve ascends ($\beta > 0$) or descends ($\beta < 0$), and the rate of change

increases as $|\beta|$ increases. When $\beta = 0$, the right-hand side of equation (2) simplifies to a constant. Then, $\pi(x)$ is identical at all x, so the curve becomes a horizontal straight line. The binary response Y is then independent of X. Figure 3.2 shows the S-shaped appearance of the model for $\pi(x)$, as fitted for the example in the following subsection. Since it is curved rather than a straight line, the rate of change in $\pi(x)$ per 1-unit increase in x depends on the value of x. A straight line drawn tangent to the curve at a particular x value, such as shown in Figure 3.2,



Figure 3.2 : Linear approximation to logistic regression curve

For logistic regression parameter β , that line has slope equal to $\beta \pi(x)[1 - \pi(x)]$. For instance, the line tangent to the curve at *x* for which $\pi(x) = 0.50$ has slope $\beta(0.50)(0.50) = 0.25\beta$; by contrast, when $\pi(x) = 0.90$ or 0.10, it has slope 0.09β . The slope approaches 0 as the probability approaches 1.0 or 0. The steepest slope occurs at x for which $\pi(x) = 0.50$. That x value relates to the logistic regression parameters by (1) $x = -\alpha/\beta$. This x value is sometimes called the *median effective level* and is denoted *EL*50. It represents the level at which each outcome has a 50% chance.

3.3 Logistic Regression with Retrospective Studies

Another property of logistic regression relates to situations in which the explanatory variable *X* rather than the response variable *Y* is random. This occurs with retrospective sampling designs. Sometimes such designs are used because one of the response categories occurs rarely, and a prospective study might have too few cases to enable us to estimate effects of predictors effectively. For a given sample size, effect estimates have smaller standard errors when the number of outcomes of the two types are similar than when they are very different.

Most commonly, retrospective designs are used with biomedical casecontrol studies. For samples of subjects having Y = 1 (cases) and having Y = 0 (controls), the value of X is observed. Evidence exists of an association between X and Y if the distribution of X values differs between cases and controls. For case- control studies, it is possible to estimate odds ratios but not other summary measures. Logistic regression parameters refer to odds and odds ratios. One can fit logistic regression models with data from case-control studies and estimate effects of explanatory variables. The intercept term α in the model is not meaningful, because it relates to the relative numbers of outcomes of y = 1 and y = 0. We do not estimate this, because the sample frequencies for y = 1 and y = 0 are fixed by the nature of the case–control study.

With case–control studies, it is not possible to estimate effects in binary models with link functions other than the logit. Unlike the odds ratio, the effect for the conditional distribution of X given Y does not then equal that for Y given X. This provides an important advantage of the logit link over links such as the probit. It is a major reason why logistic regression surpasses other models in popularity for biomedical studies.

Many case–control studies employ matching. Each case is matched with one or more control subjects. The controls are like the case on key characteristics such as age. The model and subsequent analysis should take the matching into account.

3.4 Inference for Logistic Regression

We have discussed how logistic regression helps describe the effects of a predictor on a binary response variable. We next present statistical inference for the model parameters, to help judge the significance and size of the effects.

Widely available software reports the maximum likelihood estimates of parameters and their standard errors. Sometimes sets of observations have the same values of predictor variables, such as when explanatory variables are discrete. Then, ML model fitting can treat the observations as the binomial counts of successes out of certain sample sizes, at the various combinations of values of the predictors. We will refer to this case as *grouped binary data* and the case in which each observation is a single binary outcome as *ungrouped binary data*. When at least one explanatory variable is continuous, binary data are naturally ungrouped.

3.4.1 Confidence Intervals for Effects

A large-sample Wald confidence interval for the parameter β in the logistic regression model, $logit[\pi(x)] = \alpha + \beta x$, is

$$\hat{\beta} \pm Z_{\frac{\alpha}{2}}(SE)$$

Exponentiating the endpoints yields an interval for e^{β} , the multiplicative effect on the odds of a 1-unit increase in *x*.

When *n* is small or fitted probabilities are mainly near 0 or 1, it is preferable to construct a confidence interval based on the likelihood-ratio test. This interval contains all the β_o values for which the likelihood-ratio test of $H_o: \beta = \beta_o$ has *P*-value > α .

3.4.2 Significance Testing

For the logistic regression model, $H_o: \beta = 0$ states that the probability of success is independent of X. Wald test statistics are simple. For large samples, $z = \hat{\beta}/SE$ has a standard normal distribution when $\beta = 0$. Refer z to the standard normal table to get a one-sided or two-sided P-value.

Equivalently, for the two-sided $H_a: \beta = 0$, $z^2 = (\hat{\beta}/SE)^2$ has a largesample chi-squared null distribution with df = 1. Although the Wald test is adequate for large samples, the likelihood-ratio test is more powerful and more reliable for sample sizes often used in practice. The test statistic compares the maximum L_o of the log-likelihood function when $\beta = 0$ to the maximum L_1 of the log-likelihood function for unrestricted β .

The test statistic, $-2(L_o - L_1)$, also has a large-sample chi-squared null distribution with df = 1.

3.4.3 Logistic regression with Categorical predictors

Similar to ordinary regression, Logistic regression can have multiple explanatory variables. Some or all of those predictors can be categorical, rather than quantitative.

Suppose a binary response *Y* has two binary predictors, *X* and *Z*. The data are then displayed in a $2 \times 2 \times 2$ contingency table.

Let x and z be two binary predictors, each take values 0 and 1 to represent

the two categories of those explanatory variable.

The model for P(Y=1),

Logit
$$|P(Y = 1)| = \alpha + \beta_1 x + \beta_2 z$$

has main effects for *x* and *z*. The variables *x* and *z* are called *indicator variables*. They indicate categories for the predictors. Indicator variables

are also called *dummy variables*. For this coding, Table 3.1 shows the logit values at the four combinations of values of the two predictors.

Table 3.1 Logits Implied by Indicator Variables in Model,

x	Z	Logit
0	0	α
1	0	$\alpha + \beta_1$
0	1	$\alpha + \beta_2$
1	1	$\alpha + \beta_1 + \beta_2$
1		1

Logit
$$|P(Y = 1)| = \alpha + \beta_1 x + \beta_2 z$$

This model assumes an absence of interaction. The effect of one factor is the same at each category of the other factor. At a fixed category *z* of *Z*, the effect on the logit of changing from x = 0 to x = 1 equals :

$$\left[\alpha + \beta_1(1) + \beta_2 z\right] - \left[\alpha + \beta_1(0) + \beta_2 z\right] = \beta_1$$

Thus, the difference between two logits equals the difference of log odds. Equivalently, that difference equals the log of the odds ratio between X and Y, at that category of Z. Thus, $\exp(\beta_1)$ equals the conditional odds ratio between X and Y. Controlling for Z, the odds of "success" at x = 1equal $\exp(\beta_1)$ times the odds of success at x = 0. This conditional odds ratio is the same at each category of Z. The lack of an interaction term implies a common value of the odds ratio for the partial tables at the two categories of *Z*. The model satisfies homogeneous association. Conditional independence exists between *X* and *Y*, controlling for *Z*, if $\beta_1 = 0$. In that case the common odds ratio equals 1. The simpler model,

$$\operatorname{Logit}\left[P\left(Y = 1\right)\right] = \alpha + \beta_2 z$$

then applies to the three-way table.

3.5 Building Logistic Regression Models

Having learned the basics of logistic regression, we now study issues relating to building a model with multiple predictors and checking its fit and strategies for model selection. After choosing a preliminary model, model checking explores possible lack of fit. In practice, large-sample methods of inference are not always appropriate.

For a given data set with a binary response we now discuss methods of how do we select a logistic regression model. The selection process becomes more challenging as the number of explanatory variables increases, because of the rapid increase in possible effects and interactions. There are two competing goals: The model should be complex enough to fit the data well, but simpler models are easier to interpret.

Most studies are designed to answer certain questions, which motivates certain terms in the model. To answer those questions, confirmatory

analyses use a restricted set of models. A study's theory about an effect may be tested by comparing models with and without that effect. In the absence of underlying theory, some studies are *exploratory* rather than *confirmatory*. Then, a search among many models may provide clues about which predictors are associated with the response and suggest questions for future research.

Data are unbalanced on Y if y = 1 occurs relatively few times or if y = 0 occurs relatively few times. This limits the number of predictors for which effects can be estimated precisely. One first guideline1 suggests there should ideally be at least 10 outcomes of each type for every predictor. For example, if y = 1 only 30 times out of n = 1000 observations, the model should have no more than about three predictors even though the overall sample size is large. This guideline is approximate. When not satisfied, software still fits the model. In practice, often the number of variables is large, sometimes even of similar magnitude as the number of observations. However, when this guideline is violated, ML estimates may be quite biased and estimates of standard errors may be poor.

For example, models with several predictors often suffer from *multicollinearity* – correlations among predictors making it seem that no one variable is important when all the others are in the model. A variable may seem to have little effect because it overlaps considerably with other predictors in the model, itself being predicted well by the other predictors.

Deleting such a redundant predictor can be helpful, for instance to reduce standard errors of other estimated effects.

As in ordinary regression, algorithms can select or delete predictors from a model in a stepwise manner. In exploratory studies, such model selection methods can be informative if used cautiously. Forward selection adds terms sequentially until further additions do not improve the fit. Backward elimination begins with a complex model and sequentially removes terms. At a given stage, it eliminates the term in the model that has the largest Pvalue in the test that its parameters equal zero. We test only the highestorder terms for each variable. It is inappropriate, for instance, to remove a main effect term if the model contains higher-order interactions involving that term. The process stops when any further deletion leads to a significantly poorer fit with either approach. For categorical predictors with more than two categories, the process should consider the entire variable at any stage rather than just individual indicator variables. Otherwise, the result depends on how you choose the baseline category for the indicator variables. We may add or drop the entire variable rather than just one of its indicators.

Variable selection methods need not yield a meaningful model. Use them with caution! When you evaluate many terms, one or two that are not truly important may look impressive merely due to chance.

In any case, statistical significance should not be the sole criterion for whether to include a term in a model. It is sensible to include a variable that is important for the purposes of the study and report its estimated effect even if it is not statistically significant. Keeping it in the model may help reduce bias in estimating effects of other predictors and may make it possible to compare results with other studies where the effect is significant (perhaps because of a larger sample size). Likewise, with a very large sample size sometimes a term might be statistically significant but not practically significant. You might then exclude it from the model because the simpler model is easier to interpret – for example, when the term is a complex interaction.

3.6 AIC, Model Selection, and the "Correct" Model

In selecting a model, one should not think that he/she has found the "correct" one. Any model is a simplification of reality. For example, we should not expect width to have an *exactly* linear effect on the logit probability of satellites. However, a simple model that fits adequately has the advantages of model parsimony. If a model has relatively little bias, describing reality well, it provides good estimates of outcome probabilities and of odds ratios that describe effects of the predictors.

Other criteria besides significance tests can help select a good model is best known as the *Akaike information criterion* (AIC). It judges a model by how close its fitted values tend to be to the true expected values, as summarized by a certain expected distance between the two. The optimal model is the one that tends to have its fitted values closest to the true outcome probabilities. This is the model that minimizes

 $AIC = -2(\log likelihood - number of parameters in model)$

The AIC penalizes a model for having many parameters. Even though a simple model is farther than a more complex model from the true relationship, for a sample the simple model may provide better estimates of the true expected values. For example, because the model

$$\text{Logit}\left[\pi(x)\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{10} x_{10} \text{ contains the model}$$

Logit $[\pi(x)] = \alpha + \beta_1 x$ as a special case, it is closer to the true relationship. If the true relationship is approximately linear, however, with sample data we would get better estimates of $\pi(x)$ by fitting the simpler model.

3.7 Summarizing Predictive Power: Classification Tables

Sometimes it is useful to summarize the predictive power of a binary regression model. One way to do this is with a *classification table*. This cross classifies the binary outcome y with a prediction of whether y = 0 or 1. The prediction is $\hat{y} = 1$ when $\hat{\pi_i} > \pi_o$ and $\hat{y} = 0$ when $\hat{\pi_i} \le \pi_o$, for some cutoff π_o . One possibility is to take $\pi_o = 0.50$. However, if a low (high) proportion of observations have y = 1, the model fit may never (always) have $\hat{\pi_i} > 0.50$, in which case one never (always) predicts $\hat{y} = 1$. Another

possibility takes π_o as the sample proportion of 1 outcomes, which is $\hat{\pi}_i$ for the model containing only an intercept term.

3.8 Model checking

For any particular logistic regression model, there is no guarantee that the model fits the data well. We next consider ways of checking the model fit.

3.8.1 Likelihood-Ratio Model Comparison Tests

One way to detect lack of fit uses a likelihood-ratio test is to compare the model with more complex ones. A more complex model might contain a nonlinear effect, such as a quadratic term to allow the effect of a predictor to change directions as its value increases. Models with multiple predictors would consider interaction terms. If more complex models do not fit better, this provides some assurance that a chosen model is adequate.

3.8.2 Goodness of Fit and the Deviance

A more general way to detect the lack of fit searches for *any* way the model fails. A goodness-of-fit test compares the model fit with the data. This approach regards the data as representing the fit of the most complex model possible – the saturated model, which has a separate parameter for each observation. Denote the working model by M. In testing the fit of M, we test whether *all* parameters that are in the saturated model but not in M equal zero. In GLM terminology, the likelihood-ratio statistic for this test is the deviance of the model .

In certain cases, this test statistic has a large-sample chi-squared null distribution. When the predictors are solely categorical, the data are summarized by counts in a contingency table. For the *ni* subjects at setting *i* of the predictors, multiplying the estimated probabilities of the two outcomes by *ni* yields estimated expected frequencies for y = 0 and y = 1. These are the *fitted values* for that setting.

Chapter 4 Data Analyses

4.1 Introduction

Available official statistics on the Palestinian society in the Palestinian Territory lack adequate data pertaining to victimization. This situation has prompted the PCBS to conduct a fourth household survey on this subject, making it possible to describe and assess certain aspects of victimization focusing on households victimized by criminal offenses.

This study is based on a household sample survey conducted during the period from 04/10/2008 until 31/12/2008. It provides basic indicators on various aspects of victimization, including households victimized by criminal offenses, type of criminal offense, tangible losses of crimes, etc. A special questionnaire was designed in accordance with UN standards and recommendations in the field of victimization statistics while taking the Palestinian particularities into account. The questionnaire covers the following items:

- 1. Type of criminal offense
- 2. Crime location
- 3. Crime reporting
- 4. Perpetrator

4.2 Data Description

4.2.1 Victims of Criminal Offenses at the Households Level

The results showed that (7.5%) of the Palestinian households in the Palestinian Territory were exposed to criminal offenses: (5.8%) of West Bank households were victimized and (10.9%) in Gaza Strip.

The results showed that the percentage of households in the Palestinian Territory that were exposed to theft (excluding vehicles theft) was (2.0%), vehicle theft or part of it was (6.1%), property damage was (1.7%), threat was (0.8%), and assault was (0.9%).

In general data revealed that the percentage of households victimized by criminal offenses in Gaza Strip is higher than that in the West Bank, except for households victimized by vehicle theft or part of it): (6.2%) in West Bank and (5.9%) in Gaza Strip, robbery or theft attempt: (0.7%) in West Bank and (0.3%) in Gaza Strip.

The results showed that (2.4%) of households in the Palestinian Territory were exposed to Israeli Soldiers or Settlers Harassment and Assault, compared with (7.1%) in 2004.

4.2.2 Victims of Criminal Offenses at the Individual Level

Type of Criminal Offense

The results showed that (33.9%) of individuals victims of criminal offenses in the Palestinian Territory were exposed to theft and attempted theft; (42.1%) of persons in West Bank and (26.3%) of persons in Gaza

Strip; threat\ assault occurred to (18.4%); (12.3%) in West Bank and (24.0%) in Gaza Strip, exposed to property damage (18.3%); (13.3%) in West Bank and (23.0%) in Gaza Strip, and Israeli Soldiers or Settlers Harassment and Assault affected (27.8%) of victimized persons; (30.1%) in West Bank and (25.6%) in Gaza Strip.

Figure 4.1: Percentage Distribution of Victimized Persons in the Palestinian Territory by Last Criminal Offense during Last 12 Months, 2008



Crime Location

The results indicated that (45.7%) of criminal offenses in the Palestinian Territory took place inside the house, (26.5%) nearby the house, (18.7%) took place in another place inside locality, and (9.1%) outside the locality. Criminal offenses occurring inside the house were higher in Gaza Strip (52.2%) than in West Bank (38.8%), and criminal offenses that took place outside the locality were lower in Gaza Strip (3.7%) compared with West

Bank (14.8%).



Figure 4.2: Percentage Distribution of Victimized Persons by Location of Last Crime and Type of Locality during Last 12 Months, 2008

Crime Reporting

The results showed that (53.0%) of persons victimized by criminal offenses in the Palestinian Territory reported the crimes, (47.0%) in West Bank and (57.6%) in Gaza Strip. The results showed that 14.2% of crime reporting underwent legal proceedings.

Figure 4.3: Percentage of Victimized Persons by Last Crime Reporting and Region during Last 12 Months, 2008



Perpetrator

The results showed that (33.2%) of criminal offenses against persons in the Palestinian Territory were committed by Israeli Soldiers or Settlers, (41.7%) in West Bank and (25.1%) in Gaza Strip. About (3.8%) of these criminal offenses were committed by one of the relatives, (6.5%) in West Bank and (1.4%) in Gaza Strip.





Physical Harm and Tangible Losses of Criminal Offenses

The percentage of criminal offenses that caused physical harm was (12.2%) in the Palestinian Territory, in West Bank (13.1%) compared with the Gaza Strip (11.5%).

The results show that (30.4%) of criminal offenses against persons in the Palestinian Territory caused tangible losses for more than 1000 Jordanian Dinars, distributed as (34.2%) in West Bank and (26.9%) in Gaza Strip. In about 72.7 of criminal offenses cases against persons in the Palestinian Territory, the victim was subjected to tangible losses, compared with (85.8%), (88.7%) and (78.6%) in the years 1996, 1999 and 2008 respectively.

Figure 4.5: Percentage Distribution of Victimized Persons by Party Prone to Tangible Losses of the Last Criminal Offense During Last 12 Months, 2008



Party Prone to Tangible Losses

Logistic regression analysis has been applied to analyze the data of this study. Logistic regression is ideal when a researcher is attempting to determine which variables predict group membership for pre-existing groups, particularly when the dependent variable is dichotomous. Further, logistic regression reveals the percent of the variance in the dependent variable accounted for by the independent variables. In addition, logistic regression can establish an hierarchy of significance for individual independent variables in the overall model, as well as explain interaction effects. There are several advantages for using logistic regression as the planned statistical analysis in the current study. First, the rigid assumptions of other forms of regression do not apply to logistic regression. For example, there is no assumption of a linear relationship between the dependent variable and the independent variables. Also, there is no assumption that the dependent variable is normally distributed in the population. In addition, there is no assumption of homogeneity of variance. Accordingly, it is not required that the dependent variable be homoscedastic for each level of the independent variables. Moreover, logistic regression does not assume that the error terms are normally distributed. Lastly, there is no requirement that independent variables be interval or unbounded.

The non-parametric version of logistic regression analysis was used in this study since all independent variables were categorical such as(region, sex,

job, owning a car, which party prone to tangible losses, place of the crime, attempting break the house and reception a threat call) and parametric tests require interval data. The dependent variable, which measured the outcome of victim versus no victim, was a discrete variable. As such, ordinary least squares regression could have been used to fit a linear probability model. However, because the linear probability model is heteroskedastic and could predict probabilities less than 1 or greater than 0, logistic regression was more appropriate to estimate the factors that predict victim.

In this study, all of the available independent variables used in building a logistic regression model and the data analysis. To examine the overall fit of the model, the model chi-square was computed. To examine the proportion of the variance in the dependent variable explained by the variance in the independent variables, the Cox and Snell *R*-squared and the Nagelkerke *R*-squared statistic were also computed. The logistic regression analysis revealed which factors should be included in the model to predict the outcome of victim versus not victim. There are many hypothesis which can tested here, The most important hypothesis states that there is a significant relation between the variable Households victimized by using the Wald statistic, which is the square of the asymptotic t-statistic

from the logistic regression analysis. In addition, a chi-square test for independence was used to test this hypothesis.

4.3 Logistic Regression Analyses

All hypothesized predictor variables were entered into a stepwise binary logistic regression model. In the initial analysis, a nine-predictor logistic model was fitted to the data to test the research hypothesis. The following nine predictor variables were used: Reasons for not reporting(cr11); Region; household job(HHocup); Is the household or any member of the household owned a car(Vs2); sex; which party prone to tangible losses(cr18); where did the crime happen(cr04); is there any thing indicate that somebody attempted to break the house(vs4); and did household or any member exposed to threat exception threat calls(vs7).

The resulting logistic model is as follows:

Logit (y) = -4.046 x1 + 2.7343 x2

Where y: Households victimized by criminal offense.

x1: family has its own car.

X2: male.

According to the model, the log of the odds of Households victimized by criminal offense being a victim was positively related to sex

(p< .05) and negatively related to family has its own car (p < .005) (See Table 4.1).

TABLE 4.1

Final Model							
Predictor	b	SEb	Odds	95%CI	Wald's	df	р
			ratio		Chi-		-
					square		
X1	-4.046	0.9569	0.175	0.0027to	17.8782	1	0.000
				0.1141			
X2	2.7343	0.8834	15.399	2.726to	9.5813	1	0.0020
				86.9775			
Test				χ2	df	р	
							-
Goodness-of-fit test				3.2246	5	0.665	
Hosmer & Lemeshow							

Summary of Logistic Regression Analysis Results

Variance explained by the model

To examine the proportion of the variance in the dependent variable explained by the variance in the independent variables, the Cox and Snell *R*-squared and Nagelkerke *R*-squared statistics were used. The Cox and Snell R2 = .211 and the Nagelkerke R2 = .478, indicating that the model explained between 21.1% and 47.8% of the variance.

Goodness of Fit Statistics

The Hosmer-Lemeshow (H-L) test is an inferential goodness of fit statistic used to assess the fit of a logistic model against actual outcomes - in this case, Households victimized by criminal offense. The test yielded a significant value [$\chi 2$ (5, N = 293) = 3.2246,

p > .05], indicating that the final model was a good fit of the data.

Statistical Tests of Individual Predictors in Overall Model

The Wald Chi-square statistic was used to evaluate the statistical significance of the individual regression coefficients. According to Table 4.1, the variables x1 and x2 were significant predictors of the outcome of an incident of Households to be victimized by criminal offense (p < .05). Based on the chi-square statistics for the two predictors in the final model, the Bs (coefficients) of those predictors are significantly different from 0. thus the null hypothesis (model) is rejected. Table 4.1 shows regression coefficients, Wald statistics, cofficiant for each of the significant predictors.

4.4 Odds Ratios

The odds ratio, as shown in Table 4.1, for the predictor of x1 (OR = 0.175) revealed that when family has its own car is almost six (1/ 0.175) times less chance than Households to be victimized by criminal offense than the family who doesn't have its own car. This means that 1 unit increase in family that has its own car increases the odds of being a Households victimized by criminal offense by a multiple of 0.175.

The odds ratio, as shown in Table 4.1, for the predictor of x2 (OR = 15.399) revealed that the male victim is almost fifteen times more than female victim. This means that a 1 unit increase in male victim increases the odds of being a Households victimized by criminal offense by a multiple of 15.399.

Validations of Predicted Probabilities

Overall, the model's percentage accuracy in classification (PAC) was 93.52%, which is an improvement over the null model's percentage accuracy in classification (PAC) of 91.47%. As may be seen in Table 4.2 the prediction of cases that were victims was less accurate than the prediction of cases that were not victims. The sensitivity level for accurately predicting victim was 64% with 16 of 25 cases accurately predicted. The specificity level was 96.27%, with 258 of 268 cases accurately predicted. The false positive rate, or the proportion of observations erroneously predicted to be victim (n = 10) over all cases predicted to be victim (n = 26), was 38.5%. Therefore, the positive predictive value (PPV) of the model was 61.5%, which is the proportion of observations correctly predicted to be victim (n = 16) over all observations predicted to be victim (n = 26). The false negative rate - the proportion of observations erroneously predicted to be not victim (n = 9) over all cases predicted to be not victim (n = 267)—was 3.37%. The negative predictive value (NPV) of the model was 96.6%, which is the proportion of observations correctly predicted to be not victim (n = 258) over all observations predicted to be not victim (n = 267).

Observed			Predicted			
			Households		Percentage	
			victimized by		Correct	
			criminal offense			
			victim	no		
				victim		
Null	Households	victim	0	25	0%	
model	victimized by	no	0	268	100%	
	criminal	victim				
	offense					
Total Percentage Correct					91.47%	
Final	Households	victim	16	9	64%	
model	victimized by	no	10	258	96.27%	
	criminal	victim				
	offense					
Total Percentage Correct					93.52%	

Accuracy of Predictions of victim Versus no victim: Final Models

Examination for Multicollinearity

Strong correlations between independent variables can result in multicollinearity in logistic regression models, which can inflate the variances of the parameter estimates. When there are small or moderate sample sizes, multicollinearity can result in lack of statistical significance of individual independent variables even when the overall model has achieved significance. To test for multicollinearity, the diagnostic statistics of Tolerance and Variance Inflation Factor (VIF) in linear regression were used. Tolerance and VIF values indicated no multicollinearity among the independent variables.

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

The current study examined eight factors as candidates for being most useful in identifying victims of crimes and their risk on a person to be victimized in the Palestinian society. Those factors include region, sex, job, owning a car, party prone to tangible losses, place of the crime, attempting to break the house and reception a threat call. The current study found statistical evidence to support only two of these factors: sex and the status of owning a car. In other words men who do not own cars are most exposed group to be victims of crime in the Palestinian society. This gives an indication that poor men are the most prone group to be victimized.

The model has been applied to predict the occurrence of persons to be victimized and succeeded in correctly predicting (64%) of people who have really fallen victims and (96%) of people who are vulnerable to crime. The general percentage of correct prediction was (93.5%).

5.2 Recommendations

 The questionnaire should be improved by PCBS to include further questions related to educational level, economic level and material status of the victims for deeper analysis.

- A database including crimes and offenses that include full records on perpetrators should be established in Palestine.
- The data should be updated regularly to facilitate tracking of crime indicators in Palestine.
- Plans to improve the security services and handling of crimes particularly among the poor people and within poor areas in Palestine should be developed.
- Further studies that take account of other new factors to be included in the questionnaire and using other statistical tests and techniques should be conducted.
- Cooperation between PCBS and the Palestinian Ministry of interiority should be established to collect frequent data for crime prevention.
- Males who belong to poor families are the most prone group in the Palestinian society to victimization; therefore, this group should be given more care and security measures and public awareness campaigns.

References

- Karmen, Andrew (2003), *Crime Victims: An Introduction to Victimology*, Wadsworth Publishing.

- Roodman, Allison (2000), A Test of a Model of Sexual Victimization: A Latent Variable Path Analysis.

-Alejandro Gaviria and Carmen Pagés. October 29 (1999), Patterns of Crime Victimization in Latin America.

- Agresti, Alan(2007), An Introduction to Categorical Data Analysis.

- Engelhardt, Bryan (2007), Crime and the labor market: A search model with optimal contracts

- Van Ness, Daniel (2009), *Crime and its victims: what we can do*, (InterVarsity Press, 1986).

- Dixie J. Koo, Karen L. Pierre University of Miami(2003), *The prevalence and risk factors for violent victimization among street recruited heroin users*.

- Marcotte, Dave and Markowit, Sara(2009), A cure for crime? Psychopharmaceuticals and crime trends.

- Hosmer, David and Lemshow, Stanley (2000), Applied Logistic Regression.

- Hentig von, Hans (1948) The Criminal and His Victim. New Haven: Yale U. Press.

- Karmen, A. (1992) Crime Victims. Pacific Grove: Brooks/Cole.

- Mendelsohn, B. (1963) "The Origin of the Doctrine of Victimology" Excerpta Criminologica 3:30

- Sebba, L. (1996). Third Parties, Victims and the Criminal Justice System. Ohio State University Press, Columbus.

- Lusignan, Richard (2007)"*Risk Assessment and Offender-Victim relationship in Juvenile Offenders*" International Journal of Offender Therapy and Comparative Criminology.

- Hickey, Eric (2006), *Victims. In Serial murderers and their victims.* Belmont, CA: Wadsworth Group.

- Miethe, Terance D. (1985), *The Myth or reality of victim involvement in crime:* A Review and comment on victim-precipitation research. Sociological focus.

- Godwin, Maurice (1998), Victim target networks as solvability factors in serial murder. Social behavioral and personality.

- Tyler, Kimberly A., Whitbeck, Les B., Hoyt, Dan R. and Cauce, Ana Mari (2004), *Risk Factors for Sexual Victimization Among Male and Female Homeless and Runaway Youth.*

- Sampson, M. (1985). Violent crime, income inequality, and regional culture. Another Look, Socialogical Focus, 18,199-208

-PCBS(2008); *Victimization Survey*, Palestinian Central Bureau of Statistics; Ramallah, Palestine.

- http://citation.allacademic.com
- http://www.researchgate.net